

An Organizing Principle for the Emergence of Life

by

Brian K. Davis

Research Foundation of Southern California, Inc.
8837 Villa La Jolla Dr., #13595
La Jolla, CA 92037

Summary. A multi-reaction system, with a shared reactant, evolves toward the autocatalytic reaction that maximally damps the thermodynamic and kinetic forces driving the system toward a stationary-state. This furnishes an organizing principle for the emergence of life from elementary molecules on the early Earth. The evidence suggests the spontaneous autocatalytic formose cycle, fuelled by a 1-carbon reactant (formaldehyde) and catalyzed by a 2-carbon product (glycoaldehyde), initiated the transition. Surface-bound biochemical cycles, driven initially by CO₂, subsequently evolved. They combined autocatalysis with pathway invariants, indicative of replication. Poly(triose-phosphate) formed by the reductive pentose-phosphate cycle was identified as the first replicator. Its replicative-form had a ladder-like structure with two polyphosphate chains crosslinked by 2-carboxy-3-keto-arabinitol (a triose pair). Crosslinked poly(triose-phosphate) sheets, with permeability related to crosslinking density, predated the membrane triose-phospholipid (bi)layer in this scenario. Poly(triose-phosphate) replication can transmit a triose sequence with gaps. Complementary H-bond pairs between counter-oriented 2-keto-pentol monomers would allow pre-RNA transmission of complex binary sequences. Thus, the first protocells likely preceded RNA. Poly(triose-phosphate) replication is chiral specific and this was linked to the prevalence of D-sugars. Nucleotide synthesis on a D-ribose-phosphate scaffold and RNA replication evidently followed, together with nucleoside-bearing cofactors, including acetyl-coenzyme A and resulting lipids. Thereafter, an RNA-dependent citrate cycle, amino acid pathways, genetic code, and proteins coevolved with a fluid mosaic protein/triose-phospholipid cell membrane.

Key words: driving forces; multi-reaction-path system; autocatalysis; pathway invariants; pre-RNA replication; poly(triose-phosphate) sheets

1. Introduction

Amino acid biosynthesis pathways conserve the number of reaction steps from their branch-point in the citrate cycle or central trunk, with sufficient accuracy for the path-distance model of genetic code formation (Davis, 1999) to account for more than fifty seemingly unrelated regularities attributed to the code since the mid-1960s (Davis, 2013a). Intermediates in amino acid synthesis also conserve, with few exceptions, a free α -carboxyl group (Davis, 2008, 2013a). Interestingly, citrate cycle and central trunk components likewise contain an invariant and a nearly invariant free terminal carboxyl group (Davis, 1999, 2013a). In addition to conserving evidence of the time-order of amino acid addition to the genetic code (Davis 1999, 2002, 2007, 2009, 2011, 2013a) more than 3.5×10^9 years ago (Schopf; 1993; Allwood et al., 2007), therefore, extant amino acid synthesis pathways likely conserved clues to the origin of the ancient pathways in central metabolism from which they branched.

Code domains of nearest-neighbor codons assigned to amino acids from the same synthesis family and conveyed by phylogenetically related tRNA species, sharing the same core structure group, provide compelling evidence that amino acid synthesis pathways were tRNA-dependent during code formation (Davis, 2008). Masking of the amino acid α -carboxyl by an attached tRNA cofactor, during formation of these pathways in the interval preceding the Last Common Ancestor (LCA), thus furnishes a mechanism for α -carboxyl retention by amino acid intermediates, following elimination of tRNA cofactors with the protein takeover of these pathways. Occurrence of a free terminal carboxyl in the earlier citrate cycle and central trunk reaction sequences, consequently, raises the possibility they too were once RNA-dependent.

Sequence identity in the conserved imprint of pre-LCA tRNA species for amino acids originating at different branch-points in central metabolism supports the proposition that RNA initially participated in the citrate cycle and central trunk (Davis, 2008, 2013a). Pre-LCA tRNA cognate with amino acids formed on pathways branching from central metabolism at pyruvate (Pyr), phosphoenolpyruvate (PEP), and 3-phosphoglycerate (3PGA) were found to have diversified from the tRNA cofactor/adaptor for asparagine² (Asn)¹, with citrate cycle oxaloacetate (OA) as precursor. Since tRNA^{Asn} and tRNA^{Gln} are mischarged by their respective diacid amino acid precursor during tRNA-dependent synthesis of the amide amino acids (Sheppard, 2007), aspartate¹ (Asp) evidently initiated tRNA-dependent synthesis of the amino acids having branch reactions at Pyr, PEP, and 3PGA. All central metabolism intermediates between these branch reactions and upstream precursor, OA, conserve a free terminal carboxyl group, consistent with masking by a tRNA cofactor.

This suggests pre-LCA tRNA-dependent amino acid synthesis pathways incorporated segments of the citrate cycle and central trunk (Davis, 2013a). Conversely, it appears RNA initially participated as a cofactor or scaffold in coordinating and catalyzing reaction sequences in central metabolism. Since each rotation of the reductive citrate cycle converts four CO₂ molecules into a new OA molecule, it is autocatalytic (Wachterhauser, 1992). In the event this pre-LCA reaction sequence took place on an RNA scaffold, a double-stranded RNA molecule, crosslinked (2'-C:2'-C) by a citrate molecule, conceivably split on reaction with acetyl-coenzyme A to form a single OA-RNA strand and another converted to a second OA-RNA strand on reaction with Pyr (Davis, 2013a). A replicating biochemical cycle results, as broadly envisioned by Orgel (2008).

¹ Superscripts denote amino acid synthesis path-distance from the branch reaction in central metabolism.

Poly(triose-phosphate) (PTP) and poly(pentose-bisphosphate) (PPBP) replication have been linked (Davis, 2012, 2013b) to the reductive pentose-phosphate cycle (RPC). Formation of successive replicating biochemical cycles thus accompanied the acquisition of life, corresponding, for definiteness, to appearance of the first self-propagating protocells. This event is attributed here to evolution from the spontaneous autocatalytic formose cycle (Butelrow, 1861; Breslow, 1959) to a series replicating molecular cycles that damped, with increasing efficiency, the physicochemical forces driving these reactions (Davis, 1996, 1998). In this reconstruction of events leading to the emergence of life, the source of cell membranes and sugar chirality are examined, in addition to origin of replication.

2. A physicochemical organizing principle for prebiotic evolution

Prebiotic evolution is credited with assembling elementary molecules on the early Earth, or elsewhere within the solar system (Crick and Orgel, 1973), into protocells equipped with a minimal set of biochemical pathways and RNA, or an RNA antecedent, enclosed by a simple membrane. This transition was either a spectacularly improbable event (Hoyle and Wickramasinghe, 1978) or, as portrayed here, far more deterministic than formerly depicted (Wächterhauser, 1992; Martin and Russell, 2007). Present evidence suggests the first protocells arose within the vicinity of a self-organized formose autocatalytic cycle, driven by a 1-carbon source (formaldehyde), well before appearance of RNA. Even with a short prebiotic interval, the question arises: what organizing principle promoted formation of the first protocell?

Restriction of the Darwin-Wallace theory of evolution to events following appearance of the first self-propagating protocell has contributed to the uncertainty surrounding the origin of life on this planet. This limitation can be traced to their theory being scaled to competitive

self-propagating (Malthus, 1793) among members of a population with different sets of heritable traits and different Malthusian fitness. Above average fitness should increase variant frequency, discounting stochastic fluctuations, possibly to the point of fixation within the population. A variant with below average fitness, conversely, faces extinction.

Population mean fitness is therefore non-decreasing during evolution, with a rate of increase related to the variance in fitness (fundamental theorem of natural selection; Fisher, 1930). 'Survival of the fittest' governs the direction of biological evolution in this top-down (population-scale) theory. Apart from the restriction noted to entities with sufficient complexity to display competitive self-propagation, the theory contains an ambiguity indicative of incompletely defined terms (Popper, 1976), as it asserts that propagation rate reflects variant fitness, even though fitness itself is solely determined by propagation rate.

Evolution under defined conditions, illustrated by changes in the frequency of Q β RNA virus midi-variants in the presence of Q β replicase (Kramer et al., 1974), led to the quantitative evaluation of the physicochemical forces driving macromolecular evolution *in silico* (Davis, 1996, 1998). Two principal molecular-scale forces were identified: (i) A kinetic force produced by different rates of reaction advancement. This force moves a reaction with multiple transition paths from a high to low effective activation free energy. In the Q β system, it shifts NTP condensation from producing RNA strands with long doubling times to those with the shortest doubling time. (ii) An initial reactant excess produces a thermodynamic force. It refers here to an excess of nucleoside-triphosphate (NTP) versus nucleoside-monophosphate (NMP) and pyrophosphate (PP). This force drives an evolving system toward thermodynamic equilibrium. Molecular evolution was shown to proceed along a path that optimally damps both forces (Davis, 1996, 1998). Other forces not directly participating in monomer condensation, such as those affecting strand annealing and, thereby, single

strand availability, can also alter the trajectory of an evolving system. A molecular-scale theory of evolution reveals novel states of coexistence between competing variants is possible (showing sequence fitness is not fixed, but responsive to changes in environment) and episodic events (associated with a switch from one thermodynamic force to another) can occur. 'Survival of the fittest' results when forces, other than the kinetic force, can be neglected. Significantly, the physicochemical forces driving evolution apply to molecular processes under prebiotic conditions. In particular, they provide an organizing principle for protocell formation from simple molecules in the early Archaean.

The reaction,



drives replication in the *in vitro* Q β system. NTP undergoes condensation via the i^{th} possible path. Each path corresponds to replication of an RNA variant, i-RNA, with a characteristic doubling time, related to the forward rate coefficient, k_i , of the condensation reaction. Under far from equilibrium conditions, the back-reaction will remain negligible. The apparent NTP condensation rate coefficient in a heterogeneous replication system is

$$\bar{k} = \sum_i k_i \frac{v_i P_i}{\sum_i v_i P_i} \quad (2)$$

a monomer weighted average, $\bar{k}$. P_i is the concentration of replicating i-RNA strands, containing v_i monomers per strand. The normalized frequency of NTP condensed into i-RNA, $\phi_i = v_i P_i / \sum_i v_i P_i$, changes at a rate

$$\wp_i' = \wp_i(k_i - \bar{k}) \quad [\text{AR}]$$

when strand separation follows replication (Davis, 1996). NTP condensation then proceeds as an autocatalytic reaction [AR]. Without strand separation, it follows standard reaction [SR] kinetics. NTP condensation in i-RNA replication then varies in frequency with time as,

$$\wp_i' = -\wp_i(k_i - \bar{k}) \quad [\text{SR}]$$

Time-variations in NTP condensation frequency thus occur in opposite directions during [AR] and [SR] modes of RNA synthesis. The corresponding mean rate coefficients, consequently, also change in opposite directions,

$$\bar{k} = \sum_i k_i \wp_i' = V(k) \geq 0 \quad [\text{AR}] \quad (\text{fundamental theorem}) \quad (3)$$

$$\bar{k} = \sum_i k_i \wp_i' = -\sum_i k_i \wp_i (k_i - \bar{k}) = -V(k) \leq 0 \quad [\text{SR}] \quad (4)$$

(inverse fundamental theorem)

In [AR] mode, RNA variants with the shortest doubling-time contribute most template strands. This accelerates NTP condensation and the mean forward rate coefficient increases with evolution (Eqn. 3). In [SR] mode, efficiently copied RNA templates are most rapidly converted to double-strand molecules, which do not replicate (Davis, 1995a,b). The mean forward rate coefficient accordingly decreases, as NTP condensation increasingly utilizes more slowly copied RNA strands (Eqn. 4). The simple [SR] in Eqn. 1 notably adopts [AR] kinetics, when coupled to replication.

Evolution toward the maximum rate of NTP condensation during [AR] mode synthesis and, conversely, toward the minimum rate in [SR] mode, follow a path that minimizes the action in the effective activation free energy and free energy transitions (Davis, 1996, 1998).

By advancing the NTP condensation reaction to the fastest available path, autocatalysis optimally damps the physicochemical forces driving evolution within a non-equilibrium system. The divergent paths of evolution in [AR] and [SR], competing for a shared reactant, furnishes an organizing principle for the evolution of protocells from elementary molecules, within a window of only 1×10^9 yr. following the formation of Earth (McGuinness, 2010).

3. Reconstructing prebiotic reaction sequences

In addition to identifying an organizing principle, reconstruction of the prebiotic reaction sequences played an essential role in formulating the proposed theory. The present reconstruction incorporates concepts developed in the course of investigations into coevolution of the genetic code with amino acid synthesis pathways (Davis, 1999, 2007, 2008, 2011, 2013a).

An invariant terminal carboxyl or phosphate in a central metabolism reaction sequence or branching pathway is attributed, in this endeavor, to initial masking by a macromolecular cofactor or scaffold. In an [SR] sequence, exemplified by amino acid synthesis, evidence from code structure and pre-LCA tRNA phylogenetics implicate masking by an attached bifunctional cofactor/adaptor tRNA molecule (Davis, 2008). This interpretation extended to a pre-LCA segment, from the citrate cycle or central trunk, upstream of the pathway branch point (Davis, 2013a). For an [AR] sequence, exemplified by the reductive citrate and pentose-phosphate cycles (Wachterhauser, 1992), an initial scaffold was implicated.

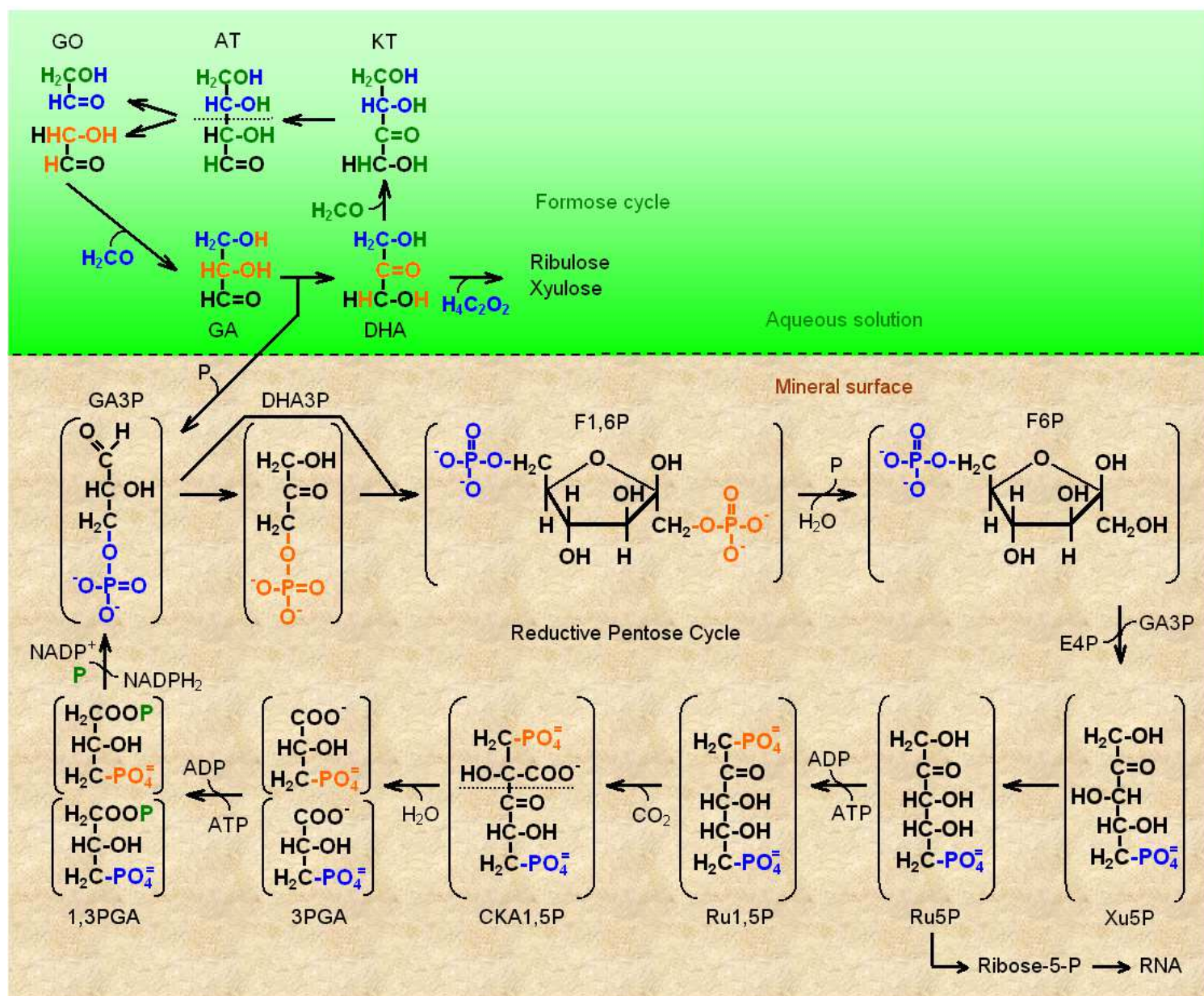
Conservation of the same free terminal group by all intermediates in a pathway conforms with the reaction sequence having remained essentially unaltered since its completion. The scope of the correlation between amino acid synthesis path length and time-order of amino acid entry into the genetic code, in accounting for code structure (Davis 2007, 2013a),

validates this assessment in relation to α -carboxyl invariance among intermediates in amino acid synthesis pathways. As these pathways coevolved with the genetic code at an advanced stage of the RNA World (Davis, 2007), they arose well after formation of the first protocells. This criterion of conservation is extended here to pathways originating in pre-RNA stages of evolution.

4. The formose cycle

Unlike the reductive pentose-phosphate cycle (RPC), the formose cycle has no invariant or charged groups (Fig. 1). Although spontaneously autocatalytic (Butelrow, 1861), it therefore shows no evidence of a scaffold required to form a replicating cycle. An absence of charged groups among formose cycle components also implies the reaction sequence took place in an aqueous medium, rather than at a mineral surface. Phosphorylated derivatives of formose cycle components glyceraldehyde (GA) and isomer dihydroxyacetone (DHA) occur

Figure 1. Invariant feature of the reductive pentose-phosphate cycle (RPC). The GA3P phosphate group, colored blue, is retained unmodified in all RPC components. This conforms with them conserving the imprint of an ancient polyphosphate scaffold, which bound RPC components to a positively charged mineral surface. Monomers in the poly(ribose-phosphate) scaffold of RNA form on cyclization of RPC component, ribulose-5-phosphate (Ru5P). Alternatively, Ru5P splits into two 3-phospho-glycerate (3PGA) molecules, following carboxylation and hydrolysis. Each triose-phosphate autocatalytically produces a new 3PGA molecule in three rotations of the RPC - fuelled by one CO₂ molecule/rotation. Poly(triose-phosphate) replication results. Glyceraldehyde (GA) and dihydroxyacetone (DHA) in the spontaneous autocatalytic formose cycle are depicted as the source of GA3P and DHA3P in the RPC. Colors mark the source of atoms in formose cycle components. No invariant charged groups occur, consistent with this simple cycle arising in an aqueous medium prior to scaffold formation. Brackets signify attachment to a polyphosphate chain. GO, glyceraldehyde; AT, aldotetrose; KT, ketotetrose; E4P, erythrose-4-phosphate; CKA1,5P, 2-carboxy-3-keto-arabinose-1,5-bisphosphate; Xu-5P, xyulose-5-phosphate, F6P, fructose-6-phosphate; F-1,6-P, fructose-1,6-bisphosphate; 1,3PGA, 1,3-bisphosphoglycerate.



in the RPC (§5). The transition to surface-bound processes is thus inferred to have followed the reaction of GA with inorganic polyphosphate strands, bound to a positively charged mineral surface, and its possible transition to DHA (Fig. 1). This may be seen as the first step toward formation of a replicating molecular cycle. Brown and Kornberg (2004) and

Achbergerova and Nahalka (2011) previously proposed polyphosphate had played a role in the origin of life.

In addition to trioses GA and DHA, the formose reaction produces tetroses, pentoses (including ribose), and keto-pentol (ribulose, xylulose) components of the RPC. A 2-carbon product of the formose cycle, glycoaldehyde (GO), catalyzes formaldehyde conversion to GA (Breslow, 1959), contributing to the robustness of the cycle during prebiotic evolution.

5. Reductive pentose-phosphate cycle as a replicating cycle

D-Ribulose-1,5 bisphosphate (RuBP) forms 2-carboxy-3-oxo-ribitol-1,5 bisphosphate (CKRuBP) or 2-carboxy-3oxo-aribinitol-1,5 bisphosphate (CKABP) on reaction with a CO_2 molecule in aqueous solution (Fig. 2). In the presence of ribulose-1,5-bisphosphate-carboxylase (RuBisCo), the latter is formed exclusively (Lorimer et al., 1986). Hydrolysis splits CKABP into two D-3-phosphoglycerate (D-3PGA) molecules. With incorporation of a single CO_2 molecule per rotation of the RPC, three RPC rotations produce a new 3PGA, making the cycle autocatalytic.

The 3PGA phosphate is retained unmodified by each RPC component. This invariance and autocatalytic nature of the RPC indicate it conserves the imprint of an early replicator. A model of the ancient replicator reveals it contained a double-strand polyphosphate scaffold, crosslinked by CKA molecules (Fig. 2).

6 Poly(triose-phosphate) replication

6.1 Replicative-form: *poly(carboxy-keto-arabinitol-1,5-bisphosphate)* ladder

RPC D-RuBP molecules form a double-strand polymer with the ladder-like structure depicted in Fig. 2. At maximum stacking density (no gaps), it has linear parallel

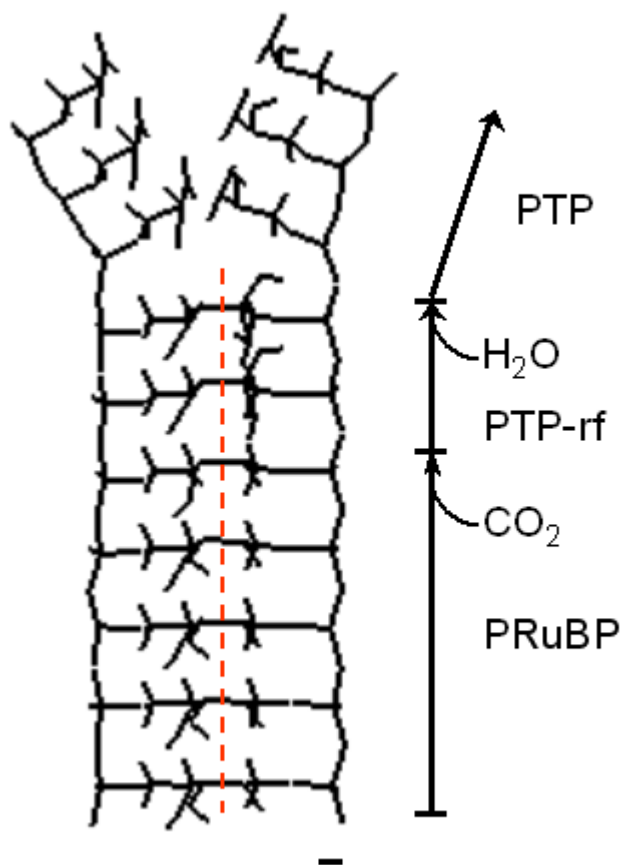


Figure 2. Stages in poly(triose-phosphate) replication. D-ribulose (Ru) crosslink parallel polyphosphate chains. A CO₂ molecule converts Ru to 2-carboxy-3-ketone-arabinitol (triose pair). Hydrolysis subsequently splits each triose pair at the pentose C2 - C3 bond (marked by orange line). Two separate poly(3-phospho-glycerate) strands result. This links the RPC cycle to a form of pre-RNA replication. The molecular model was constructed using Facio 18.1.1 (Suenaga, 2005) and optimized with restricted Hartree-Fock fragment molecular orbital calculations performed with the aid of a Gamess package (Schmidt et al., 1993). PTP, poly(triose-phosphate); PTP-rf, poly(triose-phosphate) replicative-form; PRuBP, poly(ribulose-1,5-bisphosphate). Bar length, 1Å.

polyphosphate chains, 8.8 Å apart, crosslinked by Ru molecules spaced at 2.9 Å. Reaction with CO₂ converts Ru to CKA, to yield CKABP crosslinks; CKA is preferred to epimer 2-carboxy-3-keto-ributol (CKRu), as it arises during carboxylation catalyzed by ribulose-1,5-phosphate carboxylase (RuBisCo) in the biosynthetic pathway. Hydrolysis severs the C2-C3

Ru bond, producing two poly(3PGA) strands. Since the CKA crosslinks are effectively triose pairs, poly(CKCBP) corresponds to the replicative-form of PTP².

6.2 Chiral specificity

From a space-filling CPK model of a CKABP dimer (Fig. 3), it is apparent that adjoining D- and L-isomers of coaligned CKA encounter steric hindrance in a CKABP ladder. Bulky

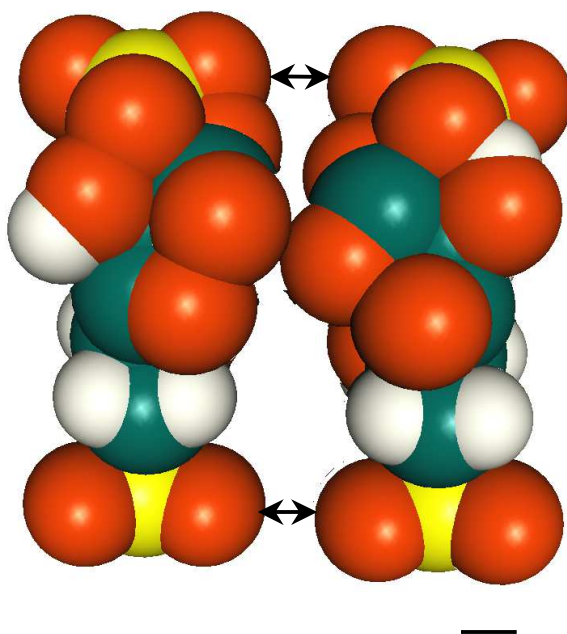


Figure 3. A space-filling CPK model showing steric hindrance between isomers of CKA1,5P. Contact between carboxylate groups of adjacent D (left) and L (right) isomers separates phosphate-group O atoms (arrows), preventing bond formation in polyphosphate chain synthesis. Yellow, P atom; green, C; orange, O; white, H. Bar length, 1 Å.

carboxyl groups prevent contact between the phosphate groups of neighboring monomers with opposite chirality. Even at low frequency, steric hindrance between neighboring

² PTP refers to poly(triose-phosphate).

monomers of mixed chirality could be prohibitive, as blocking carboxylation of adjoining and L- and D-Ru crosslinks in a poly(RuBP) ladder would halt PTP replication.

Efficient PTP replication thus requires either D- or L-monomers, but not both. The requirement for optically pure monomers in PTP replication suggests an explanation for the homochiral poly(ribose-phosphate) scaffold in RNA and its 2'-deoxy derivative DNA. Evolution driven by dissipation of the physicochemical forces within an initially weak autocatalytic reaction system could be expected to drive formation of mechanisms elevating the frequency of homochiral monomer incorporation during PTP replication (§2). Mechanisms increasing replication efficiency in other respects would be similarly enhanced during evolution. The inability of nucleotides with mixed D- and L-ribose components to form a double helix had raised the possibility that replication initially involved achiral monomers (Joyce et al., 1987), in a risk-avoidance strategy.

6.3 D-Sugar prevalence

Most cell sugars are D-isomers. In particular, the pentose-phosphate scaffold of RNA and DNA utilizes D-ribose and its 2'-deoxy derivative, respectively. Chiral selectivity of pre-LCA PTP replication, noted above, provides a mechanism to account for occurrence of D-sugars synthesized on pathways connected to the RPC. In addition, it is known that a polynucleotide double-helix can form with a scaffold having a pyranose or furanose sugar (Eschenmoser, 1999). An early reliance on cyclization of D-ribulose-5-phosphate in the RPC (Fig. 1), as sole source of the polynucleotide scaffold, reasonably accounts for the double-helix being restricted to ribose, or 2'-deoxyribose, with a furanose ring.

6.4. Poly(triose-phosphate) sheet as antecedent of the membrane phospholipid bilayer

Evidence for PTP replication places the origin of replication close to the vicinity of the self-organizing, autocatalytic formose cycle. From the residue profile of the proteolipid subunit

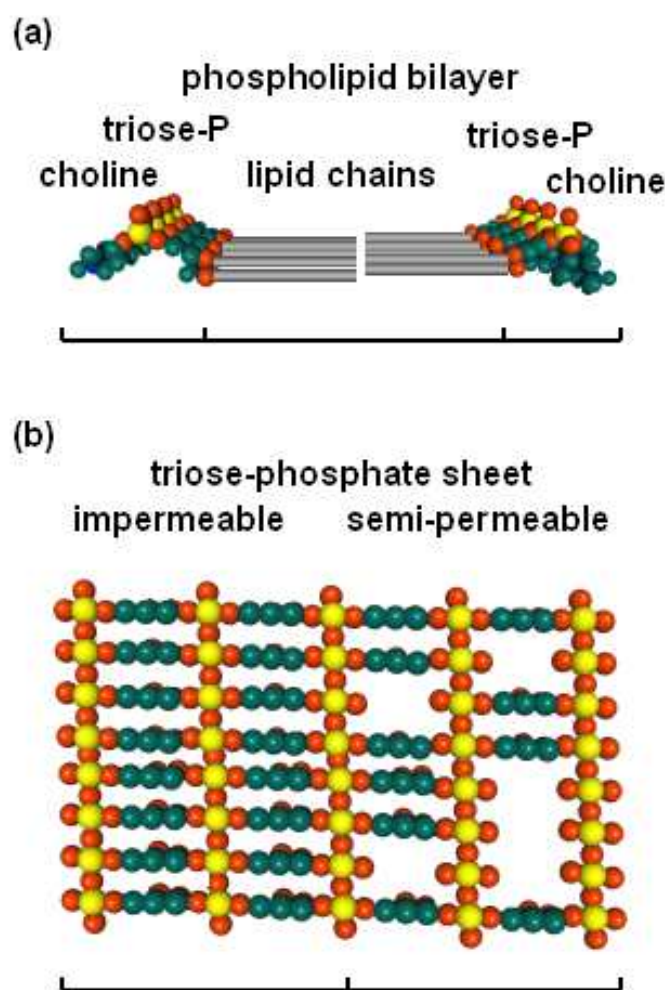


Figure 4. Comparison between cell membrane phospholipid bilayer and a poly(triose-phosphate) sheet. (a) Arrangement of triose-phosphate (glycerol-phosphate) molecules with an attached head group, exemplified by choline, and two generally unbranched fatty acid chains is depicted in the cell membrane lipid bilayer. (b) Cross-linked poly(triose-phosphate) molecules are shown to form a monomolecular sheet. The permeability of this simple membrane is portrayed as dependent on cross-linking density. Atom colors correspond to those under Fig. 3 with H atoms omitted.

of H^+ -ATPase, it is evident that proteins partitioned with the cell membrane phospholipid bilayer (phospholipid monolayer in some Archaeal species) by formation of the mid-stage genetic code (Davis, 2002). The phospholipid bilayer was apparently already formed, when membrane proteins and the 'fluid mosaic' cell membrane evolved. The dependence of phospholipid synthesis on acetyl-coenzyme A (Michal 1992) suggests membrane lipids formed after ribonucleotides. Acetyl-CoA participation in the citrate cycle also suggests membrane lipid formation was contemporaneous with evolution of the reductive citrate cycle. This CO_2 driven cycle conserves evidence of initially having had an RNA scaffold (Davis, 1999, 2013a). Since it is autocatalytic (Wächterhauser, 1992), it conforms with having originated as a biomolecular replicating cycle envisioned by Orgel (2008).

Phospholipid molecules are synthesized on a triose-phosphate (glycerol-phosphate) scaffold. Triose-phosphate sheets thus plausibly formed the first membranes, in the vicinity of the formose cycle. Figure 4a depicts fatty chains attached at C1 and C2 of L-3-glycerol-phosphate in lecithin. Triose-phospholipids self-assemble into a bilayer. Choline is attached to the polar phosphate group of glycerol-phosphate in lecithin, though serine and other attachments commonly occur. A monomolecular sheet of crosslinked triose-phosphate molecules is shown in Fig. 4b. Crosslinking density would determine the permeability of the triose-phosphate sheet.

7. Poly(ribulose-1,5-bisphosphate) ladder

7.1 Binary sequence complexity

A non-repeating pattern of triose crosslinks and gaps within a semi-permeable triose-phosphate sheet (Fig. 4b) could be replicated. Three stages in the replication of a

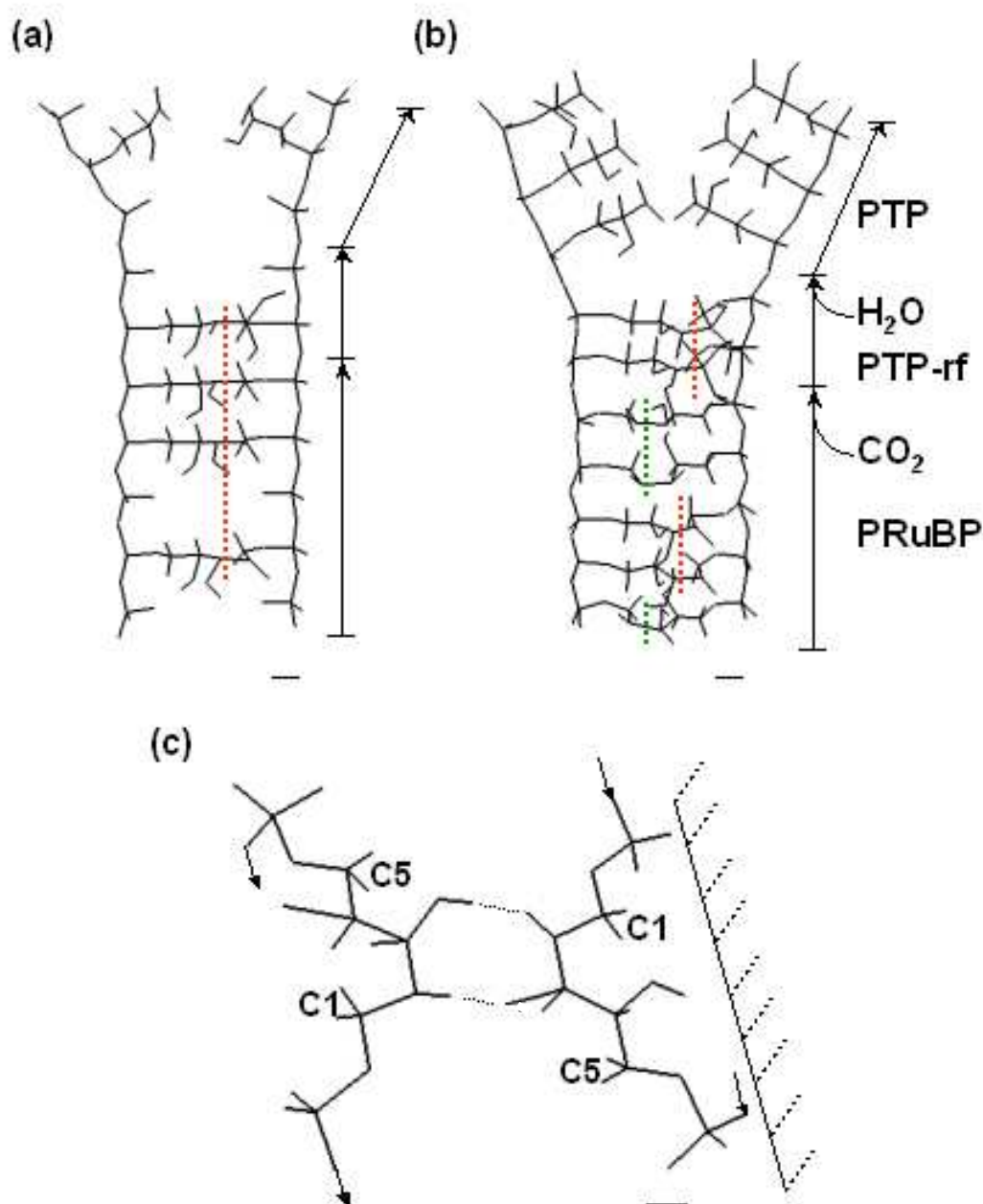


Figure 5. Sequence complexity in pre-RNA replication. (a) Shows stages in transmission of a non-repeating PTP sequence with gaps. (b) An aperiodic arrangement of ribulose monomers, oriented in a C1-C5 or C5-C1 direction between polyphosphate chains. Sequence complexity is not transmitted, as the RuBP ladder splits into two poly(3-phospho-glycerate) homopolymers. (c) The C2 ketone and C3 hydroxyl of a surface-bound RuBP, in the poly(keto-pentol-bisphosphate) molecule (KPBP), H-bond to a counter-oriented RuBP, forming a C1-C5:C5-C1 pair. Polymerization results in a daughter KPBP strand with a sequence complementary to the template. Orange line, marks C2-C3 bond in a C5 → C1 ribulose crosslink; and, green line, a C2-C3 bond in a C1 → C5 orientation. Arrows mark polyphosphate chains. Bar length, 1Å.

PTP molecule with gaps is shown in Fig. 5a. As before (Fig. 2), hydrolysis of the C2-C3 bond in CKABP produces two PTP strands, albeit each strand now shares a pattern of gaps. Synthesis of the daughter strand relies on a mechanism allowing the polyphosphate scaffold to bridge gaps between RuBP monomers. A PTP ladder with gaps results. Replication of a triose sequence with gaps corresponds to transmission of a binary sequence, which encodes the permeability of a PTP sheet.

Keto-pentose, ribulose, has a C=O at C2 and C-OH at C3 (Fig. 1). A molecular asymmetry results. Crosslinks between polyphosphate chains, in the scaffold of a poly(RuBP) ladder, are therefore directional. C1 → C5 or C5 → C1 crosslinks reverse the order of ketone and hydroxyl groups (Fig. 5c) and formation of a complex binary sequence becomes possible. Replication by splitting the poly(RuBP) ladder, however, yields two homologous PTP strands, erasing sequence complexity (Fig. 5b).

Transmission of sequence complexity in a poly(RuBP) molecule could be achieved, in principle, by replicating the poly(pentose-phosphate) ladder, rather than its PTP strands. Potential H-bonding sites, at C2 and C3, are seen to be elevated in a surface-bound poly(RuBP) ladder, attached through its anionic polyphosphate chains to a positively charged mineral surface. Counter-oriented RuBP monomers may consequently H-bond, as depicted in Fig. 5c. On polymerization of its scaffold, a daughter poly(RuBP) ladder with a complementary sequence to template would result.

Replication of the poly(RuBP) ladder, by a mechanism based on complementary H-bond formation, would provide an intermediate in the evolution of replication. As poly(RuBP) is placed between duplication of simple PTP strands, by breaking labile C-C bonds, and the copying of complex quaternary sequences in an RNA duplex, with two H-bonded

complementary nucleotide pairs. It would also illuminate the significance of the RPC ketopentol in the emergence of life.

7.2 Conformers

Poly(RuBP) conformers shown in Fig. 6 have polyphosphate chains separated by 8.8 and

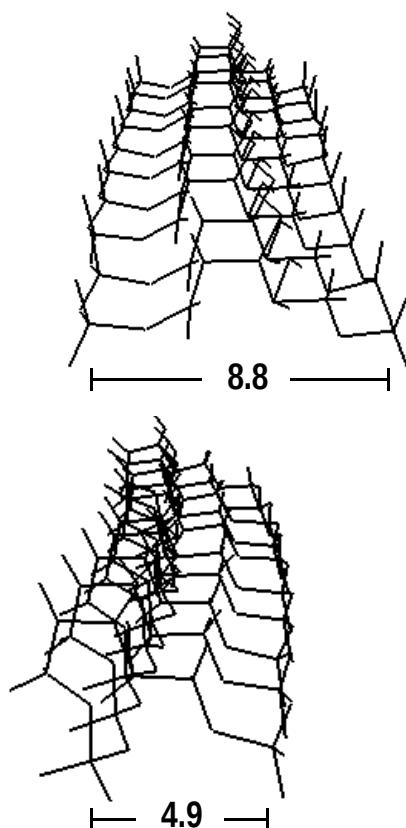


Figure 6. Extended and contracted conformers of poly(D-ribulose-1,5-bisphosphate). Inter-strand distance between polyphosphate chains is shown to decrease from 8.8 to 4.9 Å , following a counter-clockwise rotation of the C1-C2 bond, in forming O-C1-C2 angle of the lower 10-mer. Both poly(RuBP) conformers are oriented to horizontally align their polyphosphate chains.

by 4.9 Å. Rotation of the O-C1 bond about C1-C2 in opposite directions produced this difference. In the contracted conformer, O-C1 was moved counter-clockwise to form an O-C1-C2 angle of 107.8 degrees. A clockwise rotation formed the minor O-C1-C2 angle in the extended conformer. With the C1 atom as a hinge point, the transition from contracted to extended state would follow synchronous movement of polyphosphates, on the moving side of the ladder, away from a surface-bound stationary side. A nano-scale movement results. The versatility of poly(RuBP) becomes apparent from the magnitude of the effect produced by this single change to its structure.

8. The RNA scaffold

An RNA double-helix (A form) is dissected in Fig. 7 into its core of nitrogen-containing nucleoside base pairs and double-strand scaffold of D-ribose molecules linked by a phosphate group, through 3'- and 5'-C atoms.

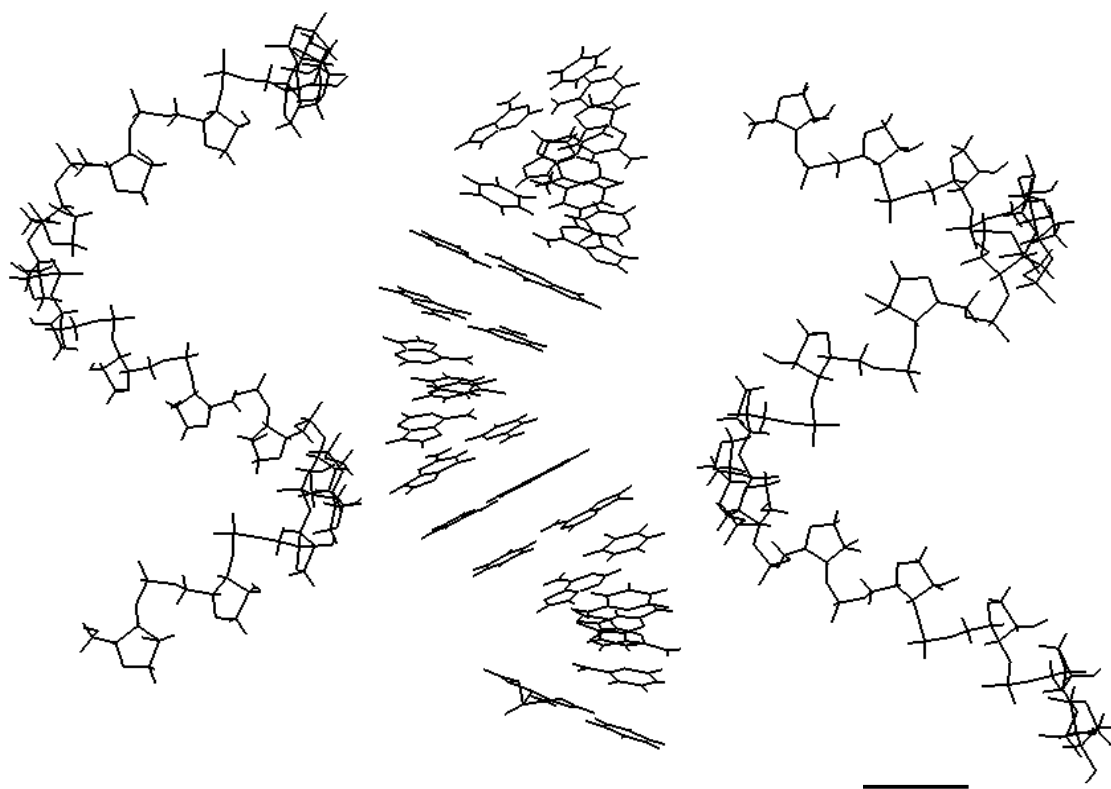


Figure 7. An RNA double-helix (A-form) separated into a double-strand poly(ribose-phosphate) scaffold and aperiodic sequence of H-bonded nucleoside base pairs. Scaffold invariance during RNA evolution necessitates formation of the poly(ribose-phosphate) chains occurred before the base sequence. The dissected RNA molecule was constructed from the Type-A DNA in the Facio 18.1.1 package (Suenaga, 2005). Bar length, 5 Å.

9.1 Antiquity

Double-helix core and scaffold contrast markedly in their evolution following species divergence from the Last Common Ancestor (LCA). Post-divergence evolution targeted the base sequence of the nucleoside core, while the D-ribose-phosphate scaffold remained invariant; the 2'-deoxy variant of RNA was formed in the pre-LCA era, during the final stage of genetic code formation (Davis, 2002). Clearly, the poly(ribose-phosphate) scaffold evolved earlier, in the pre-LCA era. The ancient pathways of nucleotide synthesis corroborate this. They demonstrate that D-ribose-5-phosphate was an invariant throughout formation of the purine synthesis pathway and beyond orotate synthesis (Michal, 1992) in formation of the pyrimidine pathway.

9.2 Nucleotide synthesis

D-Ribose-5-phosphate invariance in nucleotide synthesis reveals that evolution of the RNA scaffold can be traced back to the cyclization of D-ribulose-5-phosphate, in a branch reaction from the RPC. A search for invariants in the RPC uncovered evidence of early, pre-RNA forms of replication (§6, 7).

9. Discussion

A single molecular organizing principle, portraying evolution as a damping response to physicochemical forces acting on an evolving system, is proposed to have governed prebiotic and biological evolution. This made it possible to envision how a spontaneous autocatalytic reaction cycle, initiated by a 1-carbon molecule, led to appearance of the first protocells, equipped with interlinked molecular replication cycles. The subsequent divergence of species into separate domains and emergence of multicellular organisms are then seen as extensions of the same physicochemical process. In addition to unifying prebiotic and biological stages of evolution, the proposed theory expanded the scope of evolution dynamics beyond “survival of the fittest”, apparent from population-scale investigations (Darwin, 1859; Wallace, 1859). Since multiple forces can act on an evolving system and modify the trajectory of its evolution, occurrence of stable and unstable states of coexistence between competing species, as well as episodic events, also become possible (Davis, 1996, 1998, 2002).

To a large extent, these forces originate in the environment. The diverse set of forces the Earth environment supplied to drive the physicochemical processes occurring within each evolving system has broadly shaped the various directions taken by evolution. An abundance of water on Earth, for instance, provided a virtually limitless free energy source to drive the spread of oxygenic cyanobacteria, no longer constrained by a limited supply of H_2S (Jones, 1958). Earlier, elevated CO_2 levels on the early Earth were harnessed to drive the RPC and reductive citrate cycle. An interaction between the forces driving replication of a Q β variant, during an *in vitro* evolution experiment, and the forces promoting conversion of single strand RNA to non-replicating duplexes changed its propagation rate over time (Davis, 1995, 2000). Contrary to the fundamental theorem of natural selection (Fisher,

1930), mean fitness decreased during evolution under these circumstances. It is apparent that fitness is not a fixed attribute of a monomer sequence, or genome, as other forces in the 'environment' can mute the forces responsible for driving propagation.

Notwithstanding its significance for astrochemistry, abiotic synthesis of biogenic molecules on Earth and in space (Miller and Orgel, 1974) does not provide a credible scenario for the origin of life. Self-organizing elementary molecules in a primordial molecular into complex replicating entities requires a series of force-driven events, involving multiple transitional structures. Thermodynamic and kinetic forces vanish at equilibrium, however, making any sustained set of changes impossible in this state. The aggregation of autotrophic prokaryote species close to the root of the tree of life (Woese et al., 1990) illustrates this. By extracting free energy from a proton gradient, or other disequilibrium, they assemble simple reactants into complex biomolecules.

Application of the proposed organizing principle to the emergence of life became plausible, when the ancient pathways of central metabolism were found to contain invariants, consistent with pre-LCA formation of a series of interlinked biomolecular replicating cycles (Davis, 2012, 2013a). The significance of pathway invariants was apparent from an investigation of genetic code and amino acid synthesis coevolution (Davis 1999, 2007, 2013a). A virtually invariant α -carboxyl among amino acid intermediates corroborated evidence from code structure and pre-LCA tRNA phylogenetics that it had been masked by an attached tRNA cofactor, during formation of these pathways more than 3.5×10^9 years ago (Schopf, 1993; McGuinness, 2010). The replicating biomolecular cycles appear to have originated from the spontaneous autocatalytic formose cycle (§3, 4). It was linked directly to formation of the RPC, which conserved evidence of PTP replication (§5). Thus, protocells encapsulated by a monomolecular sheet of crosslinked poly(triose-

phosphate) strands (§6.4) conceivably arose close to the formose cycle. This limits prebiotic evolution to a pre-RNA interval. Identifying a putative intermediate between PTP duplication and RNA replication, involving complementary H-bond recognition and copying of a complex monomer sequence in a poly(RuBP) ladder, gives further support to protocells preceding RNA formation (§7).

Cyclization of D-ribulose-5-phosphate to D-ribose-5-phosphate directly links the RPC to formation of the poly(ribose-5-phosphate) scaffold in ribonucleotide synthesis and RNA replication. Coupling the ribonucleotide synthesis pathway with RNA replication fits with early evolution as a progressive series of interlinked self-propagating processes. In the evolution of replication, polyphosphate served as scaffold in the replication of a poly(sugar-phosphate), which then served as the scaffold in polynucleotide replication. Incremental increases in molecular complexity thus accompanied the advancement of replication, in a process where the replicator of one generation becomes the scaffold of the next. Restriction of RNA to a D-ribose-phosphate scaffold (Joyce et al., 1987) is attributed to the chiral selectivity of PTP replication (§6.2). As a consequence, optically pure ribose should confer a significant advantage in competitive replication. Since stereochemistry does not account for the restriction of RNA to ribose with a furanose ring (Eschenmoser, 1999), it can reasonably be linked to an early reliance on a single source of phosphorylated sugar - cyclization of ribulose-5-phosphate. Glycoaldehyde, a 2-carbon formose cycle component (Breslow, 1959), catalyzes the conversion of formaldehyde to glyceraldehyde. It may be anticipated that triozymes, tetrazymes, and pentozyms, aided by metal cofactors, will be established as catalysts in the PTP and poly(RuBP) World.

References

- Achbergerova L, Nahalka J. 2011 Polyphosphate an ancient energy source and active metabolic regulator. *Microbial Cell Factories* **10**: 63. (doi: <http://dx.doi.org/10.1186/1475-2859-10-63>)
- Allwood A.C, Walter MR, Burch IW, Kamber BS. 2007 3.43 billion-year-old stromatolite reef from the Pilbara Craton of Western Australia: Ecosystem-scale insights to early life on Earth *Precambrian Res.* **158**, 198-227.
- Breslow R. 1959 The mechanism of the formose reaction. *Tetrahedron Letters* **21**, 22-26. (doi: [http://dx.doi.org/10.1016/S0040-4039\(01\)99487-0](http://dx.doi.org/10.1016/S0040-4039(01)99487-0))
- Brown MR, Kornberg A. 2004 Inorganic polyphosphate in the origin and survival of species. *Proc. Natl. Acad. Sci USA* **101** 16085-16087. (doi: <http://dx.doi.org/10.1073/pnas.0406909101>)
- Butelrow A. 1861 Bildung einer zuckerartigen Substanz durch Synthese. *Liebigs Ann. Chem.* **120**, 295-298.
- Crick FHC, Orgel LE. 1973 Directed panspermia. *Icarus* **19**, 341-346.
- Darwin C. 1859 On the variation of organic beings in a state of nature; on the natural means of selection on the comparison of domestic races and true species. *J. Proc. Linnean Soc. (Zoology)* **3**, 46-50.
- Davis BK. 1978 Rate of polymer formation and entropy production during competitive replication. *J. Mol. Evol.* **10**, 325-338. (doi: <http://dx.doi.org/10.1007/BF01734222>)
- Davis BK. 1995a Stability of replicative form and fitness among RNA variants transcribed by Q β replicase. *Proc. Roy. Soc. London B* **260**, 39-43. (doi: <http://dx.doi.org/10.1098/rspb.1995.0056>)
- Davis BK. 1995b Significance of strand configuration in self-replicating RNA molecules. *Phil.*

- Soc Roy. Soc. London* **B350**, 345-352. (doi: <http://dx.doi.org/10.1098/rstb.1995.0169>)
- Davis BK. 1996 A theory of evolution that includes prebiotic self-organization and episodic species formation. *Bull. Math. Biol.* **58**, 65-97.
(doi: <http://dx.doi.org/10.1007/BF02458282>)
- Davis BK. 1998 The forces driving molecular evolution. *Prog. Biophys. Mol. Biol.* **69**, 83-150.
(doi: [http://dx.doi.org/10.1016/S0079-6107\(97\)00034-5](http://dx.doi.org/10.1016/S0079-6107(97)00034-5))
- Davis BK. 1999 Evolution of the genetic code. *Prog. Biophys. Mol. Biol.* **72**, 157-243.
(doi: [http://dx.doi.org/10.1016/S0079-6107\(99\)00006-1](http://dx.doi.org/10.1016/S0079-6107(99)00006-1))
- Davis, BK. 2000 Darwinian aspects of molecular evolution at sublinear propagation rates. *Bull. Math. Biol.* **62**, 675-694. (doi: <http://dx.doi.org/10.1006/bulm.2000.0173>)
- Davis BK. 2002 Molecular evolution before the origin of species. *Prog. Biophys. Mol. Biol.* **79**, 77-133. (doi: [http://dx.doi.org/10.1016/S0079-6107\(02\)00012-3](http://dx.doi.org/10.1016/S0079-6107(02)00012-3))
- Davis BK. 2007 Making sense of the genetic code with the path-distance model. In *Leading-Edge Messenger RNA Research Communications* (ed. MH. Ostrovsky) pp. 1-32. New York: Nova Science
- Davis BK. 2008 Imprint of early tRNA diversification on the genetic code: Domains of contiguous codons read by related adaptors for sibling amino acids. In *Messenger RNA Research Perspectives* (ed. T. Takayama) pp. 35-79. New York: Nova Science.
- Davis BK. 2009 On mapping the genetic code. *J.Theor. Biol.* **259**, 860-862.
(doi: <http://dx.doi.org/10.1016/j.jtbi.2009.05.009>)
- Davis BK. 2011 Genetic code domains conserve the imprint of tRNA cofactors encoded to specify cognate amino acid synthesis.
(url: <http://www.archive.org/details/GeneticCodeDomains>)
- Davis BK. 2012 Replicative-form of poly(triose-phosphate).

url:<http://www.archive.org/details/ Replicative-formOfPolytriose-phosphate>)

Davis BK 2013a Making sense of the genetic code with a path-distance model based on tRNA-dependent pathways. url: archive.org/details/MakingSenseOfGeneticCode

Davis BK. 2013b Complementary H-bond pairing mechanism for replication of a binary polypentose sequence. url: <http://archive.org/details/PolypentoseReplication>

Fisher RA. 1930 *The Genetical Theory of Natural Selection*. 2nd Ed. 1958 New York: Dover

Hoyle F, Wickramasinghe C. 1978 *Lifecloud. The Origin of Life in the Universe*. London: Dent.

Jones CW. 1985 The evolution of bacterial respiration. In *Evolution in Prokaryotes* eds.

Schleifer KH, Stackerand E. New York: Academic Press, pp. 175-204.

Joyce GF, Schwartz AW, Miller SL, Orgel LE. 1987 The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc. Natl. Acad. Sci USA* **84**, 4398-4402. (doi: <http://dx.doi.org/10.1073/pnas.84.13.4398>)

Kramer FR, Mills DR, Cole PE, Nishara T, Spiegelman S. 1974 Evolution *in vitro*: sequence and phenotype of a mutant RNA resistant to ethidium bromide. *J. Mol. Biol.* **89**, 719-738. (doi: [http://dx.doi.org/10.1016/0022-2836\(74\)90047-3](http://dx.doi.org/10.1016/0022-2836(74)90047-3))

Lorimer GH, Andrews TJ, Pierce J, Schloss JV. 1986 2'-Carboxy-3-keto-D-arabinitol-1,5-bisphosphate, the 6-carbbon intermediate of the ribulose bisphosphate carboxylase. *Phil. Trans. Roy. Soc. London* **B 313** 397-407. (doi: <http://dx.doi.org/10.1098/rstb.1986.0046>)

Malthus T. 1793 *An Essay on the Principle of Population*. J. Johnson. Reprinted 1998 Electronic Scholarly Publishing Project. url: <http://www.esp.org>

Martin W, Russell MJ. 2007 On the origin of biochemistry at an alkaline hydrothermal vent. *Phil. Trans. Roy. Soc.* **B 362**, 1887-1925. (doi: <http://dx.doi.org/10.1098/rstb.2006.1881>)

- McGuinness E. 2010 Some molecular moments of the Hadean and Archaean aeons: retrospective overview from the interfacing years of the second and third millennia. *Chem. Rev.* **110**, 5191-5215. (doi: <http://dx.doi.org/10.1021/cr050061l>)
- Michal, G. 1992 *Biochemical Pathways*. 3rd Edition, Penzberg: Boehringer Mannheim
- Miller S, Orgel LE 1974 *The Origin of Life on the Earth* Englewood Cliffs: Prentice-Hall
- Orgel, LE. 2008 The implausibility of metabolic cycles on the prebiotic Earth. *PLoS Biol* **6(1)**, e18. (doi: [10.1371/journal.pbio.0060018](http://dx.doi.org/10.1371/journal.pbio.0060018))
- Popper K. 1976 Darwinism as a metaphysical research program. In *Unended Quest: an intellectual autobiography*. London: Fontana, pp.167-180.
- Schmidt MW, Baldrige KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL, Dupuis M, Montgomery JA. 1993 Gamess general atomic and molecular electronic structure system. *J. Comput. Chem.* **14**, 1347-1363. (doi: <http://dx.doi.org/10.1002/jcc.540141112>)
- Schopf JW 1993 Microfossils of the early Archaean apex chert: new evidence of the antiquity of life. *Science* **260**, 640-646.
(doi: <http://dx.doi.org/10.1126/science.260.5108.640>)
- Sheppard K. 2007 *RNA-dependent Biosynthesis of Glutamine in Bacteria and Archaea*. Thesis. Yale Univ., New Haven.
- Suenaga M. 2005 Facio: New computational chemistry environment for PC Gamess. *J. Comput. Chem. Japan* **4**, 25-32. (doi: <http://dx.doi.org/10.2477/jccj.4.25>)
- Wächterhauser G. 1992 Groundworks for an evolutionary biochemistry: the iron-sulphur world. *Prog. Biophys. Mol. Biol.* **58**, 85-201.
(doi: [http://dx.doi.org/10.1016/0079-6107\(92\)90022-X](http://dx.doi.org/10.1016/0079-6107(92)90022-X))
- Wallace A. 1859 On the tendency of varieties to depart indefinitely from the original type. *J.*

Proc. Linnean Soc. (Zoology) **3**, 53-62.

Woese CR, Kandler O, Wheelis ML. 1990 Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eukarya. *Proc. Natl. Acad. Sci. USA* **87**, 4576-4581. (doi: <http://dx.doi.org/10.1073/pnas.87.12.4576>)